Seminar showcases Toxicogenetics Challenge winners

By Ernie Hood

Some 60 scientists gathered Dec. 2 at NIEHS for an informative seminar, *(http://www.niehs.nih.gov/funding/assets/docs/toxicogenetics_challenge_seminar_508.pdf)* "Crowdsourcing Tox21 Qualitative High Throughput Screening Data," focusing on the NIEHS-NCATS-UNC DREAM Toxicogenetics Challenge.

Speakers described the Tox21 project that generated the data used in the Challenge, the innovative approaches employed by the teams that won the two subchallenges, and the value of crowdsourcing to advance and accelerate scientific knowledge.

The June 11-Sept. 15 Challenge was co-sponsored by NIEHS, the National Center for Advancing Translational Sciences (NCATS), *(http://www.ncats.nih.gov/)* the Carolina Center for Computational Toxicology *(http://comptox.unc.edu/)* (CCCT) at the University of North Carolina at Chapel Hill (UNC), DREAM *(http://www.the-dream-project.org/)* (Dialogue for Reverse Engineering Assessments and Methods), and Sage Bionetworks. *(http://sagebase.org/)*
NIEHS and NTP Director Linda Birnbaum, Ph.D., moderated the seminar. "The Challenge represents a groundbreaking new direction for toxicity testing," she said.

The challenge data source

Nour Abdo, a doctoral student in the Department of Environmental Sciences and Engineering at UNC, spoke about a Tox21 project known as the 1000 genomes toxicity screening project, which utilizes the power of human genome variation for population-scale *in vitro* testing.

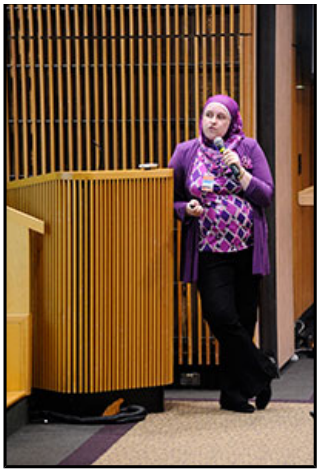
The bigger picture of challenges - the strength of crowdsourcing

DREAM representative Gustavo Stolovitzky, Ph.D., from the IBM Computational Biology Center, called into the seminar to discuss "Data Re-use and the Wisdom of Crowds." He provided a history of crowdsourcing in computational biology, and explained why it makes sense to conduct challenges and leverage the wisdom of crowds to advance biomedical research. He pointed out that data sets multiply their impact by becoming accessible to a wide segment of the scientific community, and noted that after a competition is complete, the crowd's wisdom can be further tapped by fostering collaborative research.

The project provided the raw data supplied to the Challenge competitors. Cytotoxicity data was generated on 1,086 human lymphoblast cell lines, representing nine populations from five continents, in an assay with 179 environmental chemicals at eight concentrations. The study generated roughly 2.6 million data points - the largest-scale experiment to date, in terms of the number of cell lines used and chemicals screened.

As Abdo explained, the population-based approach appears to be far more powerful than traditional *in vitro* toxicity testing. "The current toxicity testing approaches usually include model systems with very homogeneous genetics, so only the hazard can be evaluated. Our approach incorporates genetic diversity, allowing testing of not only hazards, but also the variability in the population," she said.

The population-based approach allows quantitative assessment of both hazard and interindividual variability in chemical toxicity, as well as identification of susceptible sub-populations, new understanding of the genetic determinants of interindividual variability, and generation of testable hypotheses about toxicity pathways. The data can also be used to build predictive *in silico* models.



Abdo presented data from the 1000 genomes toxicity screening project. She said that the current paradigm in toxicity screening is based on using in vivo data to derive a human reference dose for individual chemicals, but the future paradigm being developed will use in vitro data for human reference dose determination. (Photo courtesy of Steve McCaw)

The challenge winners

The challenge included two subchallenges. Subchallenge 1, which generated 99 submissions from 34 teams, involved using the supplied data to accurately predict individual responses to compound exposure, based on genomic information. Subchallenge 2, with 85 submissions from 24 teams, called for development of a model to accurately predict how a particular population would respond to certain types of chemicals. Teams from the same institution, the Quantitative Biomedical Research Center (QBRC) *(http://qbrc.swmed.edu/)* at the University of Texas Southwestern Medical Center, were named best performer in both of the subchallenges.

Among the many attentive listeners at the challenge seminar were, from right, NTP Associate Director John Bucher, Ph.D., and Health Scientist Administrator Jerry Heindel, Ph.D. (Photo courtesy of Steve McCaw)

Hao Tang, Ph.D., lead scientist for the subchallenge 2 team and a member of the subchallenge 1 team, described the methods the teams employed successfully to come in first in both segments of the competition. In subchallenge 1, the key to robust toxicity predictivity in the model was the use of fine mapping based on the geographic area where a cell line originated. For subchallenge 2, the model was enhanced, by incorporating supervised feature selection and multimodal distribution-based analysis into standard quantitative structure-activity relationship modeling.

"Those approaches, along with others contributed by many of the other challenge teams, will undoubtedly improve our ability to use both genomic data and chemical data to accurately predict cytotoxicity *in vitro*," said NTP Biomolecular Screening Branch Chief Raymond Tice, Ph.D. "The Challenge accomplished exactly what we had hoped it would."

Joined at the podium by Birnbaum, Tang, right, was clearly delighted with the response to her presentation. (Photo courtesy of Steve McCaw)

Results of the challenge will be published in the journal Nature Biotechnology.

(Ernie Hood is a contract writer with the NIEHS Office of Communications and Public Liaison.)

Ivan Rusyn, M.D., Ph.D., also from CCCT, elaborated on a point made by Tang, as he responded to a question from an attendee. "To me, as a toxicologist, one of the most important outcomes of these challenges was that it's not necessarily that we need to collect more data, but what we can do with the data that we have." (Photo courtesy of Steve McCaw)

Stolovitzky said that the wisdom of crowds means that the aggregate results from multiteam competitions are more robust and more predictive than any individual results. (Photo courtesy of Gustavo Stolovitzky)

Biostatistician Fred Wright, Ph.D., from CCCT, said the 1000 genomes toxicity screening project has shown that, while there is much evidence of heritability in susceptibility to environmental exposures, it can be difficult to isolate. "What Nour has shown is that, somehow, there is signal lurking there, but it's spread out amongst a lot of small effect variants throughout the genome," he said, in response to a question from the audience. (Photo courtesy of Steve McCaw)

Environmental Genomics Group lead Douglas Bell, Ph.D., asked about the relative merits of the array data versus RNA-Seq data, as used in the challenge. (Photo courtesy of Steve McCaw)

The Environmental Factor is produced monthly by the National Institute of Environmental Health Sciences (NIEHS) *(http://www.niehs.nih.gov/)*
, Office of Communications and Public Liaison.